

Nonparametric estimation of the distribution of the autoregressive coefficient from panel random-coefficient AR(1) data

R. Leipus^{1,2}, A. Philippe^{3,4}, V. Pilipauskaitė^{2,3*}, D. Surgailis²

October 6, 2016

¹Vilnius University, Faculty of Mathematics and Informatics, Naugarduko 24, LT-03225 Vilnius, Lithuania

²Vilnius University, Institute of Mathematics and Informatics, Akademijos 4, LT-08663 Vilnius, Lithuania

³Université de Nantes, Laboratoire de Mathématiques Jean Leray, 44322 Nantes Cedex 3, France

⁴ANJA INRIA Rennes Bretagne Atlantique

Abstract

We discuss nonparametric estimation of the distribution function $G(x)$ of the autoregressive coefficient $a \in (-1, 1)$ from a panel of N random-coefficient AR(1) data, each of length n , by the empirical distribution function of lag 1 sample autocorrelations of individual AR(1) processes. Consistency and asymptotic normality of the empirical distribution function and a class of kernel density estimators is established under some regularity conditions on $G(x)$ as N and n increase to infinity. The Kolmogorov-Smirnov goodness-of-fit test for simple and composite hypotheses of Beta distributed a is discussed. A simulation study for goodness-of-fit testing compares the finite-sample performance of our nonparametric estimator to the performance of its parametric analogue discussed in [1].

Keywords: random-coefficient autoregression, empirical process, Kolmogorov-Smirnov statistic, goodness-of-fit testing, kernel density estimator, panel data

2010 MSC: 62G10, 62M10, 62G07.

1 Introduction

Panel data can describe a large population of heterogeneous units/agents which evolve over time, e.g., households, firms, industries, countries, stock market indices. In this paper we consider a panel where each individual unit evolves over time according to order-one random coefficient autoregressive model (RCAR(1)). It is well known that aggregation of specific RCAR(1) models can explain long memory phenomenon, which is often empirically observed in economic time series (see [9] for instance). More precisely, consider a panel $\{X_i(t), t = 1, \dots, n, i = 1, \dots, N\}$, where each $X_i = \{X_i(t), t \in \mathbb{Z}\}$ is an RCAR(1) process with $(0, \sigma^2)$ noise and random coefficient $a_i \in (-1, 1)$, whose autocovariance

$$EX_i(0)X_i(t) = \sigma^2 \int_{-1}^1 \frac{x^{|t|}}{1-x^2} dG(x) \quad (1.1)$$

*Corresponding author. Email address: vytaute.pilipauskaite@gmail.com.

is determined by the distribution function $G(x) = \Pr(a \leq x)$ of the autoregressive coefficient. Granger [9] showed, for a specific Beta-type distribution $G(x)$, that the *contemporaneous* aggregation of independent processes $\{X_i(t)\}$, $i = 1, \dots, N$, results in a stationary Gaussian long memory process $\{\mathcal{X}(t)\}$, i.e.,

$$N^{-1/2} \sum_{i=1}^N X_i(t) \xrightarrow{\text{fdd}} \mathcal{X}(t) \quad \text{as } N \rightarrow \infty, \quad (1.2)$$

where the autocovariance $\text{E}\mathcal{X}(0)\mathcal{X}(t) = \text{E}X_1(0)X_1(t)$ decays slowly as $t \rightarrow \infty$ so that $\sum_{t \in \mathbb{Z}} |\text{E}\mathcal{X}(0)\mathcal{X}(t)| = \infty$.

A natural statistical problem is recovering the distribution $G(x)$ (the frequency of a across the population of individual AR(1) ‘microagents’) from the aggregated sample $\{\mathcal{X}(t), t = 1, \dots, n\}$. This problem was treated in [5, 6, 12]. Some related results were obtained in [4, 10, 11]. Albeit nonparametric, the estimators in [5, 12] involve an expansion of the density $g = G'$ in an orthogonal polynomial basis and are sensitive to the choice of the tuning parameter (the number of polynomials), being limited in practice to very smooth densities g . The last difficulty in estimation of G from aggregated data is not surprising due to the fact that aggregation *per se* inflicts a considerable loss of information about the evolution of individual ‘micro-agents’.

Clearly, if the available data comprises evolutions $\{X_i(t), t = 1, \dots, n\}$, $i = 1, \dots, N$, of all N individual ‘micro-agents’ (the panel data), we may expect a much more accurate estimate of G . Robinson [15] constructed an estimator for the moments of G using sample autocovariances of X_i and derived its asymptotic properties as $N \rightarrow \infty$, whereas the length n of each sample remains fixed. Beran et al. [1] discussed estimation of two-parameter Beta densities g from panel AR(1) data using maximum likelihood estimators with unobservable a_i replaced by sample lag 1 autocorrelation coefficient of $X_i(1), \dots, X_i(n)$ (see Section 6), and derived the asymptotic normality together with some other properties of the estimators as N and n tend to infinity.

The present paper studies nonparametric estimation of G from panel random-coefficient AR(1) data using the empirical distribution function:

$$\hat{G}_{N,n}(x) := \frac{1}{N} \sum_{i=1}^N \mathbf{1}(\hat{a}_{i,n} \leq x), \quad x \in \mathbb{R}, \quad (1.3)$$

where $\hat{a}_{i,n}$ is the lag 1 sample autocorrelation coefficient of X_i , $i = 1, \dots, N$ (see (3.3) below). We also discuss kernel estimation of the density $g(x) = G'(x)$ based on smoothed version of (1.3). We assume that individual AR(1) processes X_i are driven by identically distributed shocks containing both common and idiosyncratic (independent) components. Consistency and asymptotic normality as $N, n \rightarrow \infty$ of the above estimators are derived under some regularity conditions on $G(x)$. Our results can be applied to test goodness-of-fit of the distribution $G(x)$ to a given hypothesized distribution (e.g., a Beta distribution) using the Kolmogorov-Smirnov statistic, and to construct confidence intervals for $G(x)$ or $g(x)$.

The paper is organized as follows. Section 2 obtains the rate of convergence of the sample autocorrelation coefficient $\hat{a}_{i,n}$ to a_i , in probability, the result of independent interest. Section 3 discusses the weak convergence of the empirical process in (1.3) to a generalized Brownian bridge. The Kolmogorov-Smirnov goodness-of-fit test for simple and composite hypotheses of Beta distributed a is discussed in Section 4. In Section 5 we study kernel density estimators of $g(x)$. We show that these estimates are asymptotically normally distributed and their mean integrated square error tends to zero. A simulation study of Section 6 compares the empirical performance of (1.3) and the parametric estimator of [1] to the goodness-of-fit testing for $G(x)$ under null Beta distribution. The proofs of auxiliary statements can be found in the Appendix.

In what follows, C stands for a positive constant whose precise value is unimportant and which may change from line to line. We write \rightarrow_p , \rightarrow_d , \rightarrow_{fdd} for the convergence in probability and the convergence of (finite-

dimensional) distributions respectively, whereas \Rightarrow denotes the weak convergence in the space $D[-1, 1]$ with the supremum metric.

2 Estimation of random autoregressive coefficient

Consider an RCAR(1) process

$$X(t) = aX(t-1) + \zeta(t), \quad t \in \mathbb{Z}, \quad (2.1)$$

where innovations $\{\zeta(t)\}$ admit the following decomposition:

$$\zeta(t) = b\eta(t) + c\xi(t), \quad t \in \mathbb{Z}, \quad (2.2)$$

where random sequences $\{\eta(t)\}$, $\{\xi(t)\}$ and random coefficients a, b, c satisfy the following conditions:

Assumption A₁ $\{\eta(t)\}$ are independent identically distributed (i.i.d.) random variables (r.v.s) with $E\eta(0) = 0$, $E\eta^2(0) = 1$, $E|\eta(0)|^{2p} < \infty$ for some $p > 1$.

Assumption A₂ $\{\xi(t)\}$ are i.i.d. r.v.s with $E\xi(0) = 0$, $E\xi^2(0) = 1$, $E|\xi(0)|^{2p} < \infty$ for the same p as in A₁.

Assumption A₃ b and c are possibly dependent r.v.s such that $\Pr(b^2 + c^2 > 0) = 1$ and $Eb^2 < \infty$, $Ec^2 < \infty$.

Assumption A₄ $a \in (-1, 1)$ is a r.v. with a distribution function (d.f.) $G(x) := \Pr(a \leq x)$ supported on $[-1, 1]$ and satisfying

$$E\left(\frac{1}{1-|a|}\right) = \int_{-1}^1 \frac{dG(x)}{1-|x|} < \infty. \quad (2.3)$$

Assumption A₅ a , $\{\eta(t)\}$, $\{\xi(t)\}$ and the vector $(b, c)^\top$ are mutually independent.

Remark 2.1 In the context of panel observations (see (3.1) below), $\{\eta(t)\}$ is the common component and $\{\xi(t)\}$ is the idiosyncratic component of shocks. The innovation process $\{\zeta(t)\}$ in (2.2) is i.i.d. if the coefficients b and c are nonrandom. In the general case $\{\zeta(t)\}$ is a dependent and uncorrelated stationary process with $E\zeta(0) = 0$, $E\zeta^2(0) = Eb^2 + Ec^2$, $E\zeta(0)\zeta(t) = 0$, $t \neq 0$.

Under conditions A₁–A₅, a unique strictly stationary solution of (2.1) with finite variance exists and is written as

$$X(t) = \sum_{s \leq t} a^{t-s} \zeta(s), \quad t \in \mathbb{Z}. \quad (2.4)$$

Clearly, $EX(t) = 0$ and $EX^2(t) = E\zeta^2(0)E(1-a^2)^{-1} < \infty$. Note that (2.3) is equivalent to

$$E\left(\frac{1}{1-|a|^p}\right) < \infty, \quad 1 < p \leq 2,$$

since $1 - |a| \leq 1 - |a|^p \leq 2(1 - |a|)$ for $a \in (-1, 1)$.

For an observed sample $X(1), \dots, X(n)$ from the stationary process in (2.4), define the sample mean $\bar{X}_n := n^{-1} \sum_{t=1}^n X(t)$ and the sample lag 1 autocorrelation coefficient

$$\hat{a}_n := \frac{\sum_{t=1}^{n-1} (X(t) - \bar{X}_n)(X(t+1) - \bar{X}_n)}{\sum_{t=1}^n (X(t) - \bar{X}_n)^2}. \quad (2.5)$$

Note the estimator \hat{a}_n in (2.5) does not exceed 1 a.s. in absolute value by the Cauchy-Schwarz inequality. Moreover, it is invariant to shift and scale transformations of $\{X(t)\}$ in (2.1), i.e., we can replace $\{X(t)\}$ by $\{\rho X(t) + \mu\}$ with some (unknown) $\mu \in \mathbb{R}$ and $\rho > 0$.

Proposition 2.1 Under Assumptions A_1 – A_5 , for any $0 < \gamma < 1$ and $n \geq 1$, it holds

$$\Pr(|\hat{a}_n - a| > \gamma) \leq C(n^{-(p/2) \wedge (p-1)}\gamma^{-p} + n^{-1}),$$

with $C > 0$ independent of n, γ .

PROOF. See Appendix.

Assume now that the d.f. $G(x) = \Pr(a \leq x)$ satisfies the following Hölder condition:

Assumption A_6 There exist constants $L_G > 0$ and $\varrho \in (0, 1]$ such that

$$|G(x) - G(y)| \leq L_G |x - y|^\varrho, \quad x, y \in [-1, 1]. \quad (2.6)$$

Consider the d.f. of \hat{a}_n :

$$G_n(x) := \Pr(\hat{a}_n \leq x), \quad x \in \mathbb{R}. \quad (2.7)$$

Corollary 2.2 Let Assumptions A_1 – A_6 hold. Then, as $n \rightarrow \infty$,

$$\sup_{x \in [-1, 1]} |G_n(x) - G(x)| = O(n^{-\frac{\varrho}{\varrho+p}(\frac{p}{2} \wedge (p-1))}).$$

PROOF. Denote $\delta_n := \hat{a}_n - a$. For any (nonrandom) $\gamma > 0$ from (2.6) we have

$$\sup_{x \in [-1, 1]} |G_n(x) - G(x)| = \sup_{x \in [-1, 1]} |\Pr(a + \delta_n \leq x) - \Pr(a \leq x)| \leq L_G \gamma^\varrho + \Pr(|\delta_n| > \gamma),$$

implying

$$\sup_{x \in [-1, 1]} |G_n(x) - G(x)| \leq L_G \gamma^\varrho + C(n^{-1} + n^{-(p/2) \wedge (p-1)}\gamma^{-p})$$

with $C > 0$ independent of n, γ . Then the corollary follows from Proposition 2.1 by taking $\gamma = \gamma_n = o(1)$ such that $\gamma_n^\varrho \sim n^{-(p/2) \wedge (p-1)}\gamma_n^{-p}$ and noting that the exponent $\frac{\varrho}{\varrho+p}(\frac{p}{2} \wedge (p-1)) < 1$. \square

3 Asymptotics of the empirical distribution function

Consider random-coefficient AR(1) processes $\{X_i(t)\}$, $i = 1, 2, \dots$, which are stationary solutions to

$$X_i(t) = a_i X_i(t-1) + \zeta_i(t), \quad t \in \mathbb{Z}, \quad (3.1)$$

with innovations $\{\zeta_i(t)\}$ having the same structure as in (2.2):

$$\zeta_i(t) = b_i \eta(t) + c_i \xi_i(t), \quad t \in \mathbb{Z}. \quad (3.2)$$

More precisely, we make the following assumption:

Assumption B $\{\eta(t)\}$ satisfies A_1 ; $\{\xi_i(t)\}$, $(b_i, c_i)^\top$, a_i , $i = 1, 2, \dots$, are independent copies of $\{\xi(t)\}$, $(b, c)^\top$, a , respectively, which satisfy Assumptions A_2 – A_6 . (Note that we assume A_5 for any $i = 1, 2, \dots$)

Remark 3.1 The individual processes X_i have covariance long memory if conditions (2.3) and $\int_{-1}^1 |1 - x|^{-2} dG(x) = \infty$ hold, which is compatible with Assumption B. The same is true about the limit aggregated process in (1.2) arising when the common component is absent. On the other hand, in the presence of the common component, long memory in the limit aggregated process arises when the individual processes have infinite variance and condition (2.3) fails, see [14].

Define the sample mean $\bar{X}_{i,n} := n^{-1} \sum_{t=1}^n X_i(t)$, the corresponding sample lag 1 autocorrelation coefficient

$$\hat{a}_{i,n} := \frac{\sum_{t=1}^{n-1} (X_i(t) - \bar{X}_{i,n})(X_i(t+1) - \bar{X}_{i,n})}{\sum_{t=1}^n (X_i(t) - \bar{X}_{i,n})^2}, \quad 1 \leq i \leq N \quad (3.3)$$

and the empirical d.f.

$$\hat{G}_{N,n}(x) := \frac{1}{N} \sum_{i=1}^N \mathbf{1}(\hat{a}_{i,n} \leq x), \quad x \in \mathbb{R}. \quad (3.4)$$

Recall that (3.4) is a nonparametric estimate of the d.f. $G(x) = \Pr(a_i \leq x)$ from observed panel data $\{X_i(t), t = 1, \dots, n, i = 1, \dots, N\}$. In the following theorem we show that $\hat{G}_{N,n}(x)$ is an asymptotically unbiased estimator of $G(x)$, as n and N both tend to infinity, and prove the weak convergence of the corresponding empirical process.

Theorem 3.1 *Assume the panel data model in (3.1)–(3.2). Let Assumption B hold and $N, n \rightarrow \infty$. Then*

$$\sup_{x \in [-1, 1]} |\mathbb{E} \hat{G}_{N,n}(x) - G(x)| = O(n^{-\frac{\rho}{\rho+p}(\frac{p}{2} \wedge (p-1))}). \quad (3.5)$$

If, in addition,

$$N = o(n^{\frac{2\rho}{\rho+p}(\frac{p}{2} \wedge (p-1))}), \quad (3.6)$$

then

$$N^{1/2}(\hat{G}_{N,n}(x) - G(x)) \Rightarrow W(x), \quad x \in [-1, 1], \quad (3.7)$$

where $\{W(x), x \in [-1, 1]\}$ is a continuous Gaussian process with zero mean and $\text{cov}(W(x), W(y)) = G(x \wedge y) - G(x)G(y)$.

PROOF. Note $\hat{a}_{i,n}$, $i = 1, \dots, N$, are identically distributed, in particular, $\mathbb{E} \hat{G}_{N,n}(x) = G_n(x)$ with $G_n(x)$ defined in (2.7). Hence, (3.5) follows immediately from Corollary 2.2.

To prove the second statement of the theorem, we approximate $\hat{G}_{N,n}(x)$ by the empirical d.f.

$$\hat{G}_N(x) := \frac{1}{N} \sum_{i=1}^N \mathbf{1}(a_i \leq x), \quad x \in [-1, 1]$$

of i.i.d. r.v.s a_i , $i = 1, \dots, N$. We have $N^{1/2}(\hat{G}_{N,n}(x) - G(x)) = N^{1/2}(\hat{G}_N(x) - G(x)) + D_{N,n}(x)$ with $D_{N,n}(x) := N^{1/2}(\hat{G}_{N,n}(x) - \hat{G}_N(x))$. Since A_6 guarantees the continuity of G , it holds

$$N^{1/2}(\hat{G}_N(x) - G(x)) \Rightarrow W(x), \quad x \in [-1, 1]$$

by the classical Donsker theorem. Then (3.7) follows once we prove $\sup_{x \in [-1, 1]} |D_{N,n}(x)| \rightarrow_p 0$. By definition,

$$D_{N,n}(x) = N^{-1/2} \sum_{i=1}^N (\mathbf{1}(a_i + \delta_{i,n} \leq x) - \mathbf{1}(a_i \leq x)) = D'_{N,n}(x) - D''_{N,n}(x),$$

where $\delta_{i,n} := \hat{a}_{i,n} - a_i$, $i = 1, \dots, N$, and

$$\begin{aligned} D'_{N,n}(x) &:= N^{-1/2} \sum_{i=1}^N \mathbf{1}(x < a_i \leq x - \delta_{i,n}, \delta_{i,n} \leq 0), \\ D''_{N,n}(x) &:= N^{-1/2} \sum_{i=1}^N \mathbf{1}(x - \delta_{i,n} < a_i \leq x, \delta_{i,n} > 0). \end{aligned}$$

For $\gamma > 0$ we have

$$D'_{N,n}(x) \leq N^{-1/2} \sum_{i=1}^N \mathbf{1}(x < a_i \leq x + \gamma) + N^{-1/2} \sum_{i=1}^N \mathbf{1}(|\delta_{i,n}| > \gamma) =: V'_N(x) + V''_{N,n}.$$

(Note that $V''_{N,n}$ does not depend on x .) By Proposition 2.1, we obtain

$$EV''_{N,n} = N^{-1/2} \sum_{i=1}^N \Pr(|\delta_{i,n}| > \gamma) \leq CN^{1/2}(n^{-(p/2) \wedge (p-1)})\gamma^{-p} + n^{-1},$$

which tends to 0 when γ is chosen as $\gamma^{\varrho+p} = n^{-(p/2) \wedge (p-1)} \rightarrow 0$. Next,

$$\begin{aligned} V'_N(x) &= N^{1/2}(\widehat{G}_N(x + \gamma) - \widehat{G}_N(x)) = N^{1/2}(G(x + \gamma) - G(x)) + U_N(x, x + \gamma], \\ U_N(x, x + \gamma] &:= N^{1/2}(\widehat{G}_N(x + \gamma) - G(x + \gamma)) - N^{1/2}(\widehat{G}_N(x) - G(x)). \end{aligned}$$

The above choice of $\gamma^{\varrho+p} = n^{-(p/2) \wedge (p-1)}$ implies $\sup_{x \in [-1, 1]} N^{1/2}|G(x + \gamma) - G(x)| = O(N^{1/2}\gamma^{\varrho}) = o(1)$, whereas $U_N(x, x + \gamma]$ vanishes in the uniform metric in probability (see Lemma 7.2 in Appendix). Since $D''_{N,n}(x)$ is analogous to $D'_{N,n}(x)$, this proves the theorem. \square

Remark 3.2 (3.6) implies that $n \gg N^{(\varrho+p)/\varrho p}$ asymptotically for $p \geq 2$. Note that $(\varrho + p)/\varrho p > 1$ and $\lim_{p \rightarrow \infty} (\varrho + p)/\varrho p = 1/\varrho$ for any $\varrho \in (0, 1]$. We may conclude that Theorem 3.1 as well as other results of this paper apply to *long* panels with n increasing much faster than N , except maybe for the limiting case $p = \infty$ for $\varrho = 1$. The main reason for this conclusion is that a_i need to be accurately estimated by (3.3) in order that $\widehat{G}_{N,n}(x)$ behaves similarly to the empirical d.f. $\widehat{G}_N(x)$ based on unobserved autocorrelation coefficients $a_i, 1 \leq i \leq N$.

4 Goodness-of-fit testing

Theorem 3.1 can be used for testing goodness-of-fit. In the case of *simple* hypothesis, we test the null $H_0: G = G_0$ vs. $H_1: G \neq G_0$ with G_0 being a certain hypothetical distribution satisfying the Hölder condition in (2.6). Accordingly, the corresponding Kolmogorov-Smirnov (KS) test rejecting H_0 whenever

$$N^{1/2} \sup_{x \in [-1, 1]} |\widehat{G}_{N,n}(x) - G_0(x)| > c(\omega) \quad (4.1)$$

has asymptotic size $\omega \in (0, 1)$ provided N, n, G_0 satisfy the assumptions for (3.7) in Theorem 3.1. (Here, $c(\omega)$ is the upper ω -quantile of the Kolmogorov distribution.) However, the goodness-of-fit test in (4.1) requires the knowledge of parameters of the model considered, which is not typically a very realistic situation. Below, we consider testing *composite* hypothesis using the Kolmogorov-Smirnov statistic with estimated parameters. The parameters will be estimated by the method of moments.

Write $\mu = (\mu^{(1)}, \dots, \mu^{(m)})^\top$ and $\widehat{\mu}_{N,n} = (\widehat{\mu}_{N,n}^{(1)}, \dots, \widehat{\mu}_{N,n}^{(m)})^\top$, where

$$\mu^{(u)} := \mathbb{E}a^u = \int_{-1}^1 x^u dG(x), \quad \widehat{\mu}_{N,n}^{(u)} := \frac{1}{N} \sum_{i=1}^N (\widehat{a}_{i,n})^u, \quad 1 \leq u \leq m.$$

Proposition 4.1 *Let the panel data model in (3.1)–(3.2) satisfy Assumption B with exception of Assumption A_6 . If $N = o(n^{\frac{2}{1+p}(\frac{p}{2} \wedge (p-1))})$ as $N, n \rightarrow \infty$, then*

$$N^{1/2}(\widehat{\mu}_{N,n} - \mu) \rightarrow_d \mathcal{N}(0, \Sigma), \quad \text{where } \Sigma := (\text{cov}(a^u, a^v))_{1 \leq u, v \leq m}. \quad (4.2)$$

Proof. Write

$$N^{1/2}(\hat{\mu}_{N,n} - \mu) = N^{1/2}(\hat{\mu}_{N,n} - \hat{\mu}_N) + N^{1/2}(\hat{\mu}_N - \mu),$$

where $\hat{\mu}_N := \frac{1}{N} \sum_{i=1}^N (a_i, \dots, a_i^m)^\top$. We have $N^{1/2}(\hat{\mu}_N - \mu) \rightarrow_d \mathcal{N}(0, \Sigma)$ as $N \rightarrow \infty$ by the multivariate central limit theorem. On the other hand, $N^{1/2}(\hat{\mu}_{N,n} - \hat{\mu}_N) \rightarrow_p 0$ follows from $E|\hat{a}_n^u - a^u| \leq CE|\hat{a}_n - a| \leq C(\gamma + \Pr(|\hat{a}_n - a| > \gamma))$ and Proposition 2.1 with $\gamma^{1+p} = n^{-((p-1) \wedge (p/2))}$, proving the proposition. \square

Remark 4.1 Robinson [15, Theorem 7] discussed a different estimate of μ , which was proved to be asymptotically normal for fixed n as $N \rightarrow \infty$ in contrast to ours. However, his result holds in the case of idiosyncratic innovations only and under stronger assumptions on G than in Proposition 4.1, which do not allow for long memory.

Consider testing the composite null hypothesis that G belongs to the family $\mathcal{G} = \{G_\theta, \theta = (\alpha, \beta)^\top \in (1, \infty)^2\}$ of Beta d.f.s versus an alternative $G \notin \mathcal{G}$, where

$$G_\theta(x) = \frac{1}{B(\alpha, \beta)} \int_0^x t^{\alpha-1} (1-t)^{\beta-1} dt, \quad x \in [0, 1] \quad (4.3)$$

and $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$ is Beta function. The u th moment of G_θ is given by

$$\mu^{(u)} = \int_0^1 x^u dG_\theta(x) = \prod_{r=0}^{u-1} \frac{\alpha + r}{\alpha + \beta + r}.$$

Parameters α, β can be found from the first two moments $\mu = (\mu^{(1)}, \mu^{(2)})^\top$ as

$$\alpha = \frac{\mu^{(1)}(\mu^{(1)} - \mu^{(2)})}{\mu^{(2)} - (\mu^{(1)})^2}, \quad \beta = \frac{(1 - \mu^{(1)})(\mu^{(1)} - \mu^{(2)})}{\mu^{(2)} - (\mu^{(1)})^2}. \quad (4.4)$$

The moment-based estimator $\hat{\theta}_{N,n} := (\hat{\alpha}_{N,n}, \hat{\beta}_{N,n})^\top$ of $\theta = (\alpha, \beta)^\top$ is obtained by replacing μ in (4.4) by its estimator $\hat{\mu}_{N,n}$. The consistency and asymptotic normality of this estimator follows by the Delta method from Proposition 4.1, see Corollary 4.2 below, where we need condition $\alpha, \beta > 1$ to satisfy Assumptions A₄ and A₆.

Corollary 4.2 *Let the panel data model in (3.1)–(3.2) satisfy Assumption B. Let $G = G_\theta$, $\theta = (\alpha, \beta)^\top$ be a Beta d.f. in (4.3), where $\alpha > 1$, $\beta > 1$. Let N, n increase as in (3.6) where $\varrho = 1$. Then*

$$N^{1/2}(\hat{\theta}_{N,n} - \theta) \rightarrow_d \mathcal{N}(0, \Lambda_\theta), \quad \Lambda_\theta := \Delta^{-1} \Sigma (\Delta^{-1})', \quad (4.5)$$

where Σ is the 2×2 matrix in (4.2) and

$$\Delta := \partial\mu/\partial\theta = \begin{pmatrix} \partial\mu^{(1)}/\partial\alpha & \partial\mu^{(1)}/\partial\beta \\ \partial\mu^{(2)}/\partial\alpha & \partial\mu^{(2)}/\partial\beta \end{pmatrix}.$$

Moreover, $\hat{\theta}_{N,n}$ is asymptotically linear:

$$N^{1/2}(\hat{\theta}_{N,n} - \theta) = N^{-1/2} \sum_{i=1}^N l_\theta(a_i) + o_p(1), \quad l_\theta(x) := \Delta^{-1}(x - \mu^{(1)}, x^2 - \mu^{(2)})^\top, \quad (4.6)$$

where $El_\theta(a) = \int_0^1 l_\theta(x) dG_\theta(x) = 0$ and $El_\theta(a)l_\theta(a)^\top = \int_0^1 l_\theta(x)l_\theta(x)^\top dG_\theta(x) = \Lambda_\theta$.

Corollary 4.3 *Let assumptions of Corollary 4.2 hold. Then*

$$N^{1/2}(\widehat{G}_{N,n}(x) - G_{\widehat{\theta}_{N,n}}(x)) \Rightarrow V_{\theta}(x), \quad x \in [0, 1],$$

where $\{V_{\theta}(x), x \in [0, 1]\}$ is a continuous Gaussian process with zero mean and covariance

$$\begin{aligned} \text{cov}(V_{\theta}(x), V_{\theta}(y)) &= G_{\theta}(x \wedge y) - G_{\theta}(x)G_{\theta}(y) \\ &\quad - \int_0^x l_{\theta}(u)^{\top} dG_{\theta}(u) \partial_{\theta} G_{\theta}(y) - \int_0^y l_{\theta}(u)^{\top} dG_{\theta}(u) \partial_{\theta} G_{\theta}(x) + \partial_{\theta} G_{\theta}(x)^{\top} \Lambda_{\theta} \partial_{\theta} G_{\theta}(y), \end{aligned}$$

where $\partial_{\theta} G_{\theta}(x) := \partial G_{\theta}(x) / \partial \theta = (\partial G_{\theta}(x) / \partial \alpha, \partial G_{\theta}(x) / \partial \beta)^{\top}$, $x \in [0, 1]$ and Λ_{θ} is defined in (4.5).

Proof. The d.f. G_{θ} with $\alpha > 1, \beta > 1$ satisfies Assumptions A₄ and A₆ with $\varrho = 1$. Recall $\widehat{G}_N(x) := N^{-1} \sum_{i=1}^N \mathbf{1}(a_i \leq x)$, $x \in [0, 1]$. Since condition (3.6) is satisfied, so $N^{1/2} \sup_{x \in [0, 1]} |\widehat{G}_{N,n}(x) - \widehat{G}_N(x)|$ vanishes in probability by Theorem 3.1, whereas the convergence $N^{1/2}(\widehat{G}_N(x) - G_{\widehat{\theta}_{N,n}}(x)) \Rightarrow V_{\theta}(x)$, $x \in [0, 1]$ follows from (4.6) using the fact that $\partial_{\theta} G_{\theta}(x)$, $x \in [0, 1]$ is continuous in θ , see [7] or [20, Theorem 19.23]. \square

With Corollary 4.3 in mind, the Kolmogorov-Smirnov test for the composite hypothesis $G \in \mathcal{G}$ can be defined as

$$\sup_{x \in [0, 1]} N^{1/2} |\widehat{G}_{N,n}(x) - G_{\widehat{\theta}_{N,n}}(x)| > c(\omega, \widehat{\theta}_{N,n}), \quad (4.7)$$

where $c(\omega, \theta)$ is the upper ω -quantile of the distribution of $\sup_{x \in [0, 1]} |V_{\theta}(x)|$:

$$\Pr \left(\sup_{x \in [0, 1]} |V_{\theta}(x)| > c(\omega, \theta) \right) = \omega, \quad \omega \in (0, 1).$$

The test in (4.7) has correct asymptotic size for any $\omega \in (0, 1)$, which follows from Corollary 4.3 and the continuity of the quantile function $c(\omega, \theta)$ in θ , see [19, p. 69], [20]. By writing $N^{1/2}(\widehat{G}_{N,n}(x) - G_{\widehat{\theta}_{N,n}}(x)) = N^{1/2}(\widehat{G}_{N,n}(x) - G(x)) + N^{1/2}(G(x) - G_{\widehat{\theta}_{N,n}}(x))$, it follows that the Kolmogorov-Smirnov statistic on the l.h.s. of (4.7) tends to infinity (in probability) under any fixed alternative $G \notin \mathcal{G}$ which cannot be approximated by a Beta d.f. G_{θ} in the uniform metric, i.e., such that $\inf_{\theta} \sup_{x \in [0, 1]} |G(x) - G_{\theta}(x)| > 0$. Moreover, even under the alternative, we preserve the consistency of $\widehat{\mu}_{N,n}$, hence $c(\omega, \widehat{\theta}_{N,n})$ being a continuous function of sample moments, converges in probability to some finite limit. Therefore the test (4.7) is consistent.

In practice, the evaluation of $c(\omega, \theta)$ requires Monte Carlo approximation which is time-consuming. Alternatively, [18, 19] discussed parametric bootstrap procedures to produce asymptotically correct critical values. We note that the assumptions of [19, Theorem 1] are valid for the family of Beta d.f.s and the moment-based estimator of θ in Corollary 4.3. The consistency of the test when using bootstrap critical values follows by a similar argument as in (4.7).

5 Kernel density estimation

In this section we assume G has a bounded probability density function $g(x) = G'(x)$, $x \in [-1, 1]$, implying Assumption A₆ with Hölder exponent $\varrho = 1$ in (2.6). It is of interest to estimate $g(x)$ in a nonparametric way from $\widehat{a}_{1,n}, \dots, \widehat{a}_{N,n}$ (3.3).

Consider the kernel density estimator

$$\widehat{g}_{N,n}(x) := \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - \widehat{a}_{i,n}}{h}\right), \quad x \in \mathbb{R}, \quad (5.1)$$

where K is a kernel, satisfying Assumption A₇ and $h = h_{N,n}$ is a bandwidth which tends to zero as N and n tend to infinity.

Assumption A₇ $K : [-1, 1] \rightarrow \mathbb{R}$ is a continuous function of bounded variation that satisfies $\int_{-1}^1 K(x)dx = 1$. Set $\|K\|_2^2 := \int_{-1}^1 K(x)^2 dx$ and $\mu_2(K) := \int_{-1}^1 x^2 K(x)dx$ and $K(x) := 0$, $x \in \mathbb{R} \setminus [-1, 1]$.

We consider two cases separately.

Case (i) $\Pr(b_1 = 0) = 1$, meaning that the coefficient $b_i = 0$ for the common shock in (3.2) is zero and that the individual processes $\{X_i(t)\}$, $i = 1, 2, \dots$, are independent and satisfy

$$X_i(t) = a_i X_i(t-1) + c_i \xi_i(t), \quad t \in \mathbb{Z}.$$

Case (ii) $\Pr(b_1 \neq 0) > 0$, meaning that $\{X_i(t)\}$, $i = 1, 2, \dots$, are mutually dependent processes.

Proposition 5.1 *Let Assumptions B and A₇ hold. If $h^{1+p} n^{(p/2) \wedge (p-1)} \rightarrow \infty$, then*

$$\mathbb{E} \hat{g}_{N,n}(x) \rightarrow g(x) \tag{5.2}$$

at every continuity point $x \in \mathbb{R}$ of g . Moreover, if

$$\begin{cases} n^{(p/2) \wedge (p-1)} h^{1+p} \rightarrow \infty & \text{in Case (i),} \\ n^{(p/2) \wedge (p-1)} (h/N)^{1+p} \rightarrow \infty & \text{in Case (ii),} \end{cases} \tag{5.3}$$

then

$$Nh \operatorname{cov}(\hat{g}_{N,n}(x_1), \hat{g}_{N,n}(x_2)) \rightarrow \begin{cases} g(x_1) \|K\|_2^2 & \text{if } x_1 = x_2, \\ 0 & \text{if } x_1 \neq x_2 \end{cases} \tag{5.4}$$

at any continuity points $x_1, x_2 \in \mathbb{R}$ of g . If $Nh \rightarrow \infty$ holds in addition to (5.3), then the estimator $\hat{g}_{N,n}(x)$ is consistent at each continuity point $x \in \mathbb{R}$:

$$\mathbb{E} |\hat{g}_{N,n}(x) - g(x)|^2 \rightarrow 0. \tag{5.5}$$

PROOF. Throughout the proof, let $K_h(x) := K(x/h)$, $x \in \mathbb{R}$. Consider (5.2). Note $\mathbb{E} \hat{g}_{N,n}(x) = h^{-1} \mathbb{E} K_h(x - \hat{a}_n)$, because $\hat{a}_{i,n}$, $i = 1, \dots, N$, are identically distributed. Let us approximate $\hat{g}_{N,n}(x)$ by

$$\hat{g}_N(x) := \frac{1}{Nh} \sum_{i=1}^N K_h(x - a_i), \quad x \in \mathbb{R}, \tag{5.6}$$

which satisfies $\mathbb{E} \hat{g}_N(x) = h^{-1} \mathbb{E} K_h(x - a) \rightarrow g(x)$ as $h \rightarrow 0$ at a continuity point x of g , see [13]. Integration by parts and Corollary 2.2 yield

$$\begin{aligned} h |\mathbb{E} \hat{g}_{N,n}(x) - \mathbb{E} \hat{g}_N(x)| &= \left| \int_{\mathbb{R}} (G_n(y) - G(y)) dK_h(x - y) \right| \\ &\leq V(K) \sup_{y \in [-1, 1]} |G_n(y) - G(y)| = O(n^{-(p/2) \wedge (p-1)/(1+p)}), \end{aligned} \tag{5.7}$$

uniformly in $x \in \mathbb{R}$, where $V(K)$ denotes the total variation of K and $V(K) = V(K_h)$. This proves (5.2).

Next, let us prove (5.4). We have

$$Nh \operatorname{cov}(\widehat{g}_N(x_1), \widehat{g}_N(x_2)) = \frac{1}{h} \mathbb{E} K_h(x_1 - a) K_h(x_2 - a) \rightarrow \begin{cases} g(x_1) \|K\|_2^2 & \text{if } x_1 = x_2, \\ 0 & \text{if } x_1 \neq x_2, \end{cases} \quad (5.8)$$

as $h \rightarrow 0$ at any points x_1, x_2 of continuity of g , see [13]. Split $Nh\{\operatorname{cov}(\widehat{g}_{N,n}(x_1), \widehat{g}_{N,n}(x_2)) - \operatorname{cov}(\widehat{g}_N(x_1), \widehat{g}_N(x_2))\} = \sum_{i=1}^3 Q_i(x_1, x_2)$, where

$$\begin{aligned} Q_1(x_1, x_2) &:= h^{-1} \{\mathbb{E} K_h(x_1 - \widehat{a}_n) K_h(x_2 - \widehat{a}_n) - \mathbb{E} K_h(x_1 - a) K_h(x_2 - a)\}, \\ Q_2(x_1, x_2) &:= h^{-1} \{\mathbb{E} K_h(x_1 - \widehat{a}_n) \mathbb{E} K_h(x_2 - \widehat{a}_n) - \mathbb{E} K_h(x_1 - a) \mathbb{E} K_h(x_2 - a)\}, \\ Q_3(x_1, x_2) &:= (N-1)h^{-1} \operatorname{cov}(K_h(x_1 - \widehat{a}_{1,n}), K_h(x_2 - \widehat{a}_{2,n})). \end{aligned}$$

Note $Q_3(x_1, x_2) = 0$ in Case (i). Similarly to (5.7),

$$|Q_1(x_1, x_2)| = h^{-1} \left| \int_{\mathbb{R}} (G_n(y) - G(y)) dK_h(x_1 - y) K_h(x_2 - y) \right| \leq C h^{-1} n^{-(p/2) \wedge (p-1)/(1+p)} \rightarrow 0$$

since $V(K_h(x_1 - \cdot) K_h(x_2 - \cdot)) \leq C$ and $|Q_2(x_1, x_2)| \leq C h^{-1} n^{-(p/2) \wedge (p-1)/(1+p)} \rightarrow 0$ uniformly in x_1, x_2 . Finally,

$$\begin{aligned} |Q_3(x_1, x_2)| &= \frac{N-1}{h} \left| \int_{\mathbb{R}} \int_{\mathbb{R}} (\Pr(\widehat{a}_{1,n} \leq y_1, \widehat{a}_{2,n} \leq y_2) - \Pr(\widehat{a}_{1,n} \leq y_1) \Pr(\widehat{a}_{2,n} \leq y_2)) dK_h(x_1 - y_1) dK_h(x_2 - y_2) \right| \\ &\leq \frac{CN}{h} \sup_{y_1, y_2 \in [-1, 1]} |\Pr(\widehat{a}_{1,n} \leq y_1, \widehat{a}_{2,n} \leq y_2) - \Pr(\widehat{a}_{1,n} \leq y_1) \Pr(\widehat{a}_{2,n} \leq y_2)| \\ &= O(Nh^{-1} n^{-(p/2) \wedge (p-1)/(1+p)}) = o(1), \end{aligned}$$

proving (5.4) and the proposition. \square

Remark 5.1 It follows from the proof of the above proposition that in the case of a (uniformly) continuous density $g(x), x \in [-1, 1]$, relations (5.2), (5.5) and the first relation in (5.4) hold uniformly in $x \in \mathbb{R}$, implying the convergence of the mean integrated squared error:

$$\int_{-\infty}^{\infty} \mathbb{E} |\widehat{g}_{N,n}(x) - g(x)|^2 dx \rightarrow 0.$$

Proposition 5.2 (Asymptotic normality) *Let Assumptions B and A_7 hold and assume $Nh \rightarrow \infty$ in addition to (5.3). Moreover, let K be a Lipschitz function in Case (ii). Then*

$$\frac{\widehat{g}_{N,n}(x) - \mathbb{E} \widehat{g}_{N,n}(x)}{\sqrt{\operatorname{var}(\widehat{g}_{N,n}(x))}} \rightarrow_d \mathcal{N}(0, 1) \quad (5.9)$$

at every continuity point $x \in (-1, 1)$ of g .

PROOF. First, consider Case (i). Since $\widehat{g}_{N,n}(x) = (Nh)^{-1} \sum_{i=1}^N V_{i,N}$ is a (normalized) sum of i.i.d. r.v.s $V_{i,N} := K_h(x - \widehat{a}_{i,n})$ with common distribution $V_N := V_{1,N}$, it suffices to verify Lyapunov's condition

$$\frac{\mathbb{E} |V_N - \mathbb{E} V_N|^{2+\delta}}{N^{\delta/2} \{\operatorname{var}(V_N)\}^{(2+\delta)/2}} \rightarrow 0, \quad (5.10)$$

for some $\delta > 0$. This follows by the same arguments as in [13]. Analogously to Proposition 5.1, we have $E|V_N|^{2+\delta} = E|K_h(x - \hat{a}_n)|^{2+\delta} \sim hg(x) \int_{-1}^1 |K(y)|^{2+\delta} dy = O(h)$ while $\text{var}(V_N) = Nh^2 \text{var}(\hat{g}_{N,n}(x)) \sim hg(x) \|K\|_2^2$ according to (5.4). Hence the l.h.s. of (5.10) is $O((Nh)^{-\delta/2}) = o(1)$, proving (5.9) in Case (i).

Let us turn to Case (ii). It suffices to prove that $\sqrt{Nh}(\hat{g}_{N,n}(x) - \hat{g}_N(x)) \rightarrow_p 0$, for $\hat{g}_N(x)$ given in (5.6). By $|K(x) - K(y)| \leq L_K|x - y|$, $x, y \in \mathbb{R}$, for $\epsilon > 0$

$$\begin{aligned} \Pr\left(\sqrt{Nh}|\hat{g}_{N,n}(x) - \hat{g}_N(x)| > \epsilon\right) &\leq \Pr\left(\frac{L_K}{\sqrt{Nh}} \sum_{i=1}^N \frac{|\hat{a}_{i,n} - a_i|}{h} > \epsilon\right) \\ &\leq N \Pr\left(|\hat{a}_n - a| > \sqrt{Nh} \left(\frac{h}{N}\right) \frac{\epsilon}{L_K}\right) \\ &\leq C\left(h(Nh)^{-p/2} \left(\frac{N}{h}\right)^{1+p} n^{-(p/2) \wedge (p-1)} + \frac{N}{n}\right) = o(1) \end{aligned}$$

from Proposition 2.1 and (5.3) with $Nh \rightarrow \infty$. □

Corollary 5.3 *Let assumptions of Proposition 5.2 hold with $h \sim cN^{-1/5}$ for some $c > 0$, i.e.,*

$$N = \begin{cases} o(n^{\frac{5}{3} \frac{1}{1+p} (\frac{p}{2} \wedge (p-1))}) & \text{in Case (i),} \\ o(n^{\frac{5}{6} \frac{1}{1+p} (\frac{p}{2} \wedge (p-1))}) & \text{in Case (ii).} \end{cases}$$

Moreover, let $g \in C^2[-1, 1]$ and $\int_{-1}^1 yK(y)dy = 0$. Then

$$N^{2/5}(\hat{g}_{N,n}(x) - g(x)) \rightarrow_d \mathcal{N}(\mu(x), \sigma^2(x)),$$

where $\mu(x) := (c^2/2)g''(x)\mu_2(K)$ and $\sigma^2(x) := (1/c)g(x)\|K\|_2^2$.

PROOF. This follows from Proposition 5.2, by noting that $E\hat{g}_N(x) - g(x) \sim h^2g''(x)\mu_2(K)/2$ as $h \rightarrow 0$ and $E\hat{g}_{N,n}(x) - E\hat{g}_N(x) = O(h^{-1}n^{-(p/2) \wedge (p-1)/(1+p)})$ by (5.7). □

6 Simulations

In this section we compare our nonparametric goodness-of-fit test in (4.1) for testing the null hypothesis $G = G_0$ with its parametric analogue studied in [1]. In accordance with the last paper, we assume $\{X_i(t)\}$ in (3.1) to be independent AR(1) processes with standard normal i.i.d. innovations $\{\zeta_i(t)\}$, $\zeta(0) \sim \mathcal{N}(0, 1)$ and the random autoregressive coefficient $a_i \in (0, 1)$ having a Beta-type density $g(x)$ with unknown parameters $\theta := (\alpha, \beta)^\top$:

$$g(x) = \frac{2}{B(\alpha, \beta)} x^{2\alpha-1} (1-x^2)^{\beta-1}, \quad x \in (0, 1), \quad \alpha > 1, \beta > 1. \quad (6.1)$$

Note that $\beta \in (1, 2)$ implies the long memory property in $\{X_i(t)\}$. Beran et al. [1] discuss a maximum likelihood estimate $\hat{\theta}_{N,n,\kappa} = (\hat{\alpha}, \hat{\beta})^\top$ of $\theta = (\alpha, \beta)^\top$ when each unobservable coefficient a_i is replaced by its estimate $\hat{a}_{i,n,\kappa} := \min\{\max\{\hat{a}_{i,n}, \kappa\}, 1 - \kappa\}$ with $\hat{a}_{i,n}$ given in (3.3) and $0 < \kappa = \kappa(N, n) \rightarrow 0$ is a truncation parameter. Under certain conditions on $N, n \rightarrow \infty$ and $\kappa \rightarrow 0$, Beran et al. [1, Theorem 2] showed that

$$N^{1/2}(\hat{\theta}_{N,n,\kappa} - \theta_0) \rightarrow_d \mathcal{N}(0, A^{-1}(\theta_0)), \quad (6.2)$$

where θ_0 is the true parameter vector,

$$A(\theta) := \begin{pmatrix} \psi_1(\alpha) - \psi_1(\alpha + \beta) & -\psi_1(\alpha + \beta) \\ -\psi_1(\alpha + \beta) & \psi_1(\beta) - \psi_1(\alpha + \beta) \end{pmatrix},$$

and $\psi_1(x) := d^2 \ln \Gamma(x)/dx^2$ is the Trigamma function. Based on (4.1) and (6.2), we consider testing both ways (nonparametrically and parametrically) the hypothesis that the unobserved autoregressive coefficients a_1, \dots, a_N are drawn from the reference distribution G_0 having density function in (6.1) with a specific θ_0 , i.e., the null $G = G_0$ vs. the alternative $G \neq G_0$. The respective test statistics are

$$T_1 := N^{1/2} \sup_x |\widehat{G}_{N,n}(x) - G_0(x)| \quad \text{and} \quad T_2 := N(\widehat{\theta}_{N,n,\kappa} - \theta_0)^\top A(\theta_0)(\widehat{\theta}_{N,n,\kappa} - \theta_0). \quad (6.3)$$

Under the null hypothesis, the distributions of statistics T_1 and T_2 converge to the Kolmogorov distribution and the chi-square distribution with 2 degrees of freedom, respectively, see (4.1), (6.2).

To compare the performance of the above testing procedures, we compute the empirical distribution of the p-value of T_1 and T_2 under null and alternative hypotheses. The p-value of observed T_i is defined as $p(T_i) = 1 - \mathcal{K}_i(T_i)$, $i = 1, 2$, where $\mathcal{K}_i(y)$, $i = 1, 2$ denote the limit distribution functions of (6.3). Recall that when the significance level of the test is correct, the (asymptotic) distribution of the p-value is uniform on $[0, 1]$. The simulation procedure to compare the performance of T_1 and T_2 is the following:

Step S₀ We fix the parameter under the null hypothesis $H_0 : \theta = \theta_0$ with $\theta_0 = (2, 1.4)^\top$.

Step S₁ We simulate 5000 panels with $N = 250$, $n = 817$ for five chosen values $\theta = (2, 1.2)^\top, (2, 1.3)^\top, (2, 1.4)^\top, (2, 1.5)^\top, (2, 1.6)^\top$ of Beta parameters.

Step S₂ For each simulated panel we compute the p-value of statistics T_1 and T_2 .

Step S₃ The empirical c.d.f.'s of computed p-values of statistics T_1 and T_2 are plotted.

The values of Beta parameters $\theta_0 = (2, 1.4)^\top$, N , n were chosen in accordance with the simulation study in [1].

Fig. 1 presents the simulation results under the true hypothesis $\theta = \theta_0$ with zoom-in on small p-values. We see that both c.d.f.'s in the left graph are approximately linear. Somewhat surprisingly, it appears that the empirical size of T_1 (the nonparametric test) is better than the size of T_2 (the parametric test). Particularly, for significance levels 0.05 and 0.1 we provide the empirical size values in Table 1.

Fig. 2 gives the graphs of the empirical c.d.f.'s of p-values of T_1 and T_2 for several alternatives $\theta \neq \theta_0$. It appears that for $\beta > \beta_0 = 1.4$ the parametric test T_2 is more powerful than the nonparametric test T_1 but for $\beta < \beta_0$ the power differences are less significant. Table 1 illustrates the empirical power for the significance levels 0.05, 0.1.

Signif. level	5%					10%				
β	1.2	1.3	1.4	1.5	1.6	1.2	1.3	1.4	1.5	1.6
T_1	.532	.137	.049	.208	.576	.653	.223	.103	.315	.702
T_2	.500	.104	.077	.313	.735	.634	.184	.134	.421	.827

Table 1: Numerical results of the comparison for testing procedure $H_0 : \theta = (2, 1.4)^\top$ at significance level 5% and 10% . The column for $\beta = 1.4$ provides the empirical size.

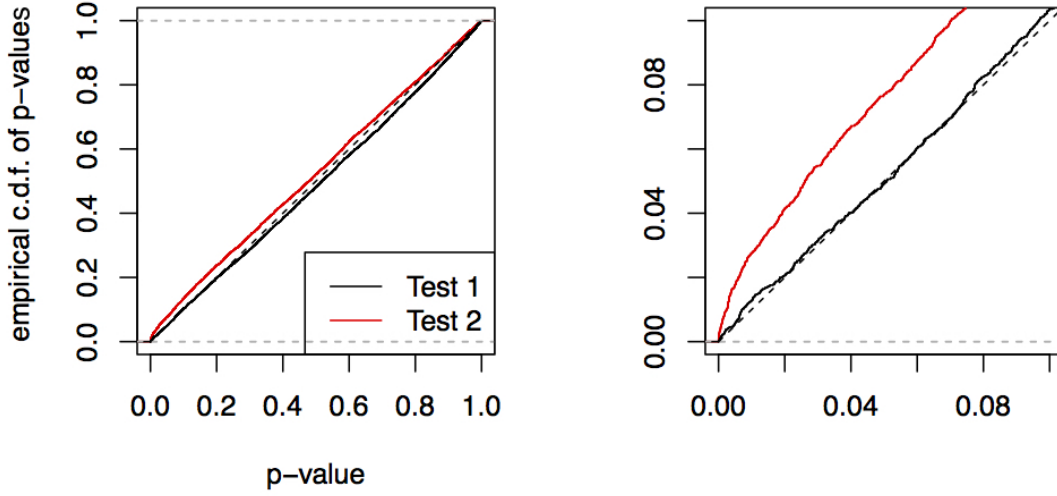


Figure 1: [left] Empirical c.d.f. of p-values of T_1 and T_2 under $H_0 : \theta_0 = (2, 1.4)^\top$; 5000 replications with $N = 250$, $n = 817$. [right] Zoom-in on the region of interest: p-values smaller than 0.1.

The above simulations (Fig. 1 and 2, Table 1) refer to the case of independent individual processes $\{X_i(t)\}$. There are no theoretical results for the parametric test T_2 , when AR(1) series are dependent. Although the nonparametric test T_1 is valid for the latter case, one may expect that the presence of the common shock component in the panel data in (3.2) has a negative effect on the test performance for short series. To illustrate this effect, we simulate 5000 panels with AR(1) processes $\{X_i(t)\}$ driven by dependent shocks in (3.2) with $b_i = b$, $c_i = (1 - b^2)^{1/2}$. As previously, we choose $\theta_0 = (2, 1.4)^\top$, $N = 250$, $n = 817$ and we fix $\theta = (2, 1.4)^\top$ to evaluate the empirical size of T_1 . Fig. 3 [left] presents the graphs of the empirical c.d.f.'s of the p-values of T_1 for $b = 1$, $b = 0.6$ and $b = 0$, the latter corresponding to independent individual processes as in Fig. 1. We see that the size of the test worsens when b increases, particularly when $b = 1$ and the individual processes are all driven by the same common noise. To overcome the last effect, the sample length n of each series in the panel may be increased as in Fig. 3 [right], where the choice of $n = 5500$ and $b = 1$ shows a much better performance of T_1 under the null hypothesis $\theta = \theta_0 = (2, 1.4)^\top$ and the alternative ($\theta = (2, 1.5)^\top$ and $\theta = (2, 1.6)^\top$) scenarios.

In conclusion,

1. We do not observe an important loss of the power for the nonparametric KS test T_1 compared to the parametric approach.
2. The KS test T_1 does not require to choose any tuning parameter contrary to the test T_2 .
3. One can use the KS test T_1 under weaker assumptions on AR(1) innovations. We only impose moment conditions. The dependence between the series is allowed by (3.2).

Acknowledgements

We thank the referees for valuable comments that led to an improved version of this paper. The first, third and fourth authors also acknowledge the support by a grant (No. MIP-063/2013) from the Research Council of Lithuania.

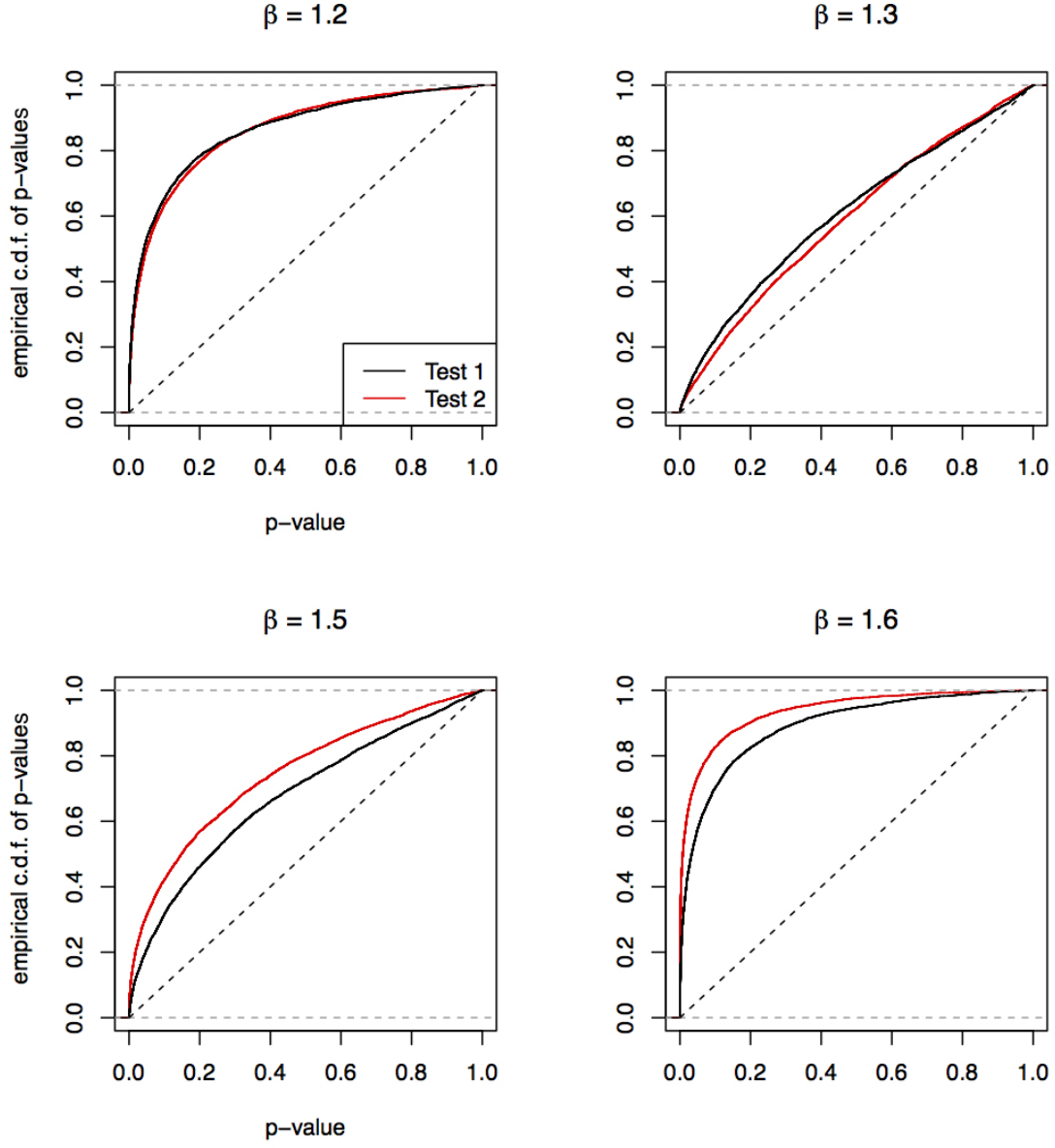


Figure 2: Empirical c.d.f. of p-values of T_1 and T_2 for testing $H_0 : \theta_0 = (2, 1.4)^\top$ under several alternatives of the form $\theta = (2, \beta)^\top$; 5000 replications with $N = 250$, $n = 817$.

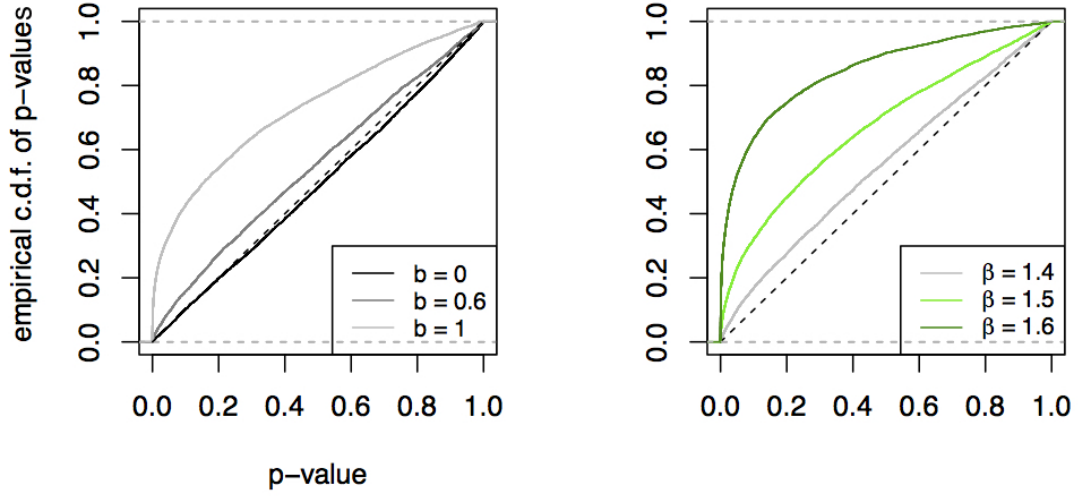


Figure 3: [left] Empirical c.d.f. of p-values of T_1 under $H_0: \theta_0 = (2, 1.4)^\top$ for different dependence structure between AR(1) series : $b_i = b$ and $c_i = \sqrt{1 - b^2}$ and $N = 250$, $n = 817$. [right] Empirical c.d.f. of p-values of T_1 for testing $H_0: \theta_0 = (2, 1.4)^\top$. AR(1) series are driven by common innovations, i.e., $b_i = 1$, $c_i = 0$, for $\theta = (2, \beta)^\top$; 5000 replications with $N = 250$, $n = 5500$.

References

- [1] J. Beran, M. Schützner, and S. Ghosh. From short to long memory: Aggregation and estimation. *Computational Statistics and Data Analysis*, 54:2432–2442, 2010.
- [2] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.
- [3] D. L. Burkholder. Distribution function inequalities for martingales. *Annals of Probability*, 1:19–42, 1973.
- [4] D. Celov, R. Leipus, and A. Philippe. Time series aggregation, disaggregation and long memory. *Lithuanian Mathematical Journal*, 47:379–393, 2007.
- [5] D. Celov, R. Leipus, and A. Philippe. Asymptotic normality of the mixture density estimator in a disaggregation scheme. *Journal of Nonparametric Statistics*, 22:425–442, 2010.
- [6] T. T. Chong. The polynomial aggregated AR(1) model. *Econometrics Journal*, 9:98–122, 2006.
- [7] J. Durbin. Weak convergence of the sample distribution function when parameters are estimated. *Annals of Statistics*, 1:279–290, 1973.
- [8] L. Giraitis, H. L. Koul, and D. Surgailis. *Large Sample Inference for Long Memory Processes*. Imperial College Press, London, 2012.
- [9] C. W. J. Granger. Long memory relationship and the aggregation of dynamic models. *Journal of Econometrics*, 14:227–238, 1980.
- [10] L. Horváth and R. Leipus. Effect of aggregation on estimators in AR(1) sequence. *TEST*, 18:546–567, 2009.
- [11] M. Jirak. Limit theorems for aggregated linear processes. *Advances in Applied Probability*, 45:520–544, 2013.
- [12] R. Leipus, G. Oppenheim, A. Philippe, and M.-C. Viano. Orthogonal series density estimation in a disaggregation scheme. *Journal of Statistical Planning and Inference*, 136:2547–2571, 2006.
- [13] E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.
- [14] D. Puplinskaitė and D. Surgailis. Aggregation of random coefficient AR(1) process with infinite variance and common innovations. *Lithuanian Mathematical Journal*, 49:446–463, 2009.
- [15] P. M. Robinson. Statistical inference for a random coefficient autoregressive model. *Scandinavian Journal of Statistics*, 5:163–168, 1978.
- [16] H. P. Rosenthal. On the subspaces of L^p ($p > 2$) spanned by sequences of independent random variables. *Israel Journal of Mathematics*, 8:273–303, 1970.
- [17] G.R. Shorack and J.A. Wellner. *Empirical Processes with Applications to Statistics*. Wiley, New York, 1986.

- [18] W. Stute, W. Gonz  les-Manteiga, and M. Presedo-Quindimil. Bootstrap based goodness-of-fit tests. *Metrika*, 40:243–256, 1993.
- [19] G. Sz  cs. Parametric bootstrap tests for continuous and discrete distributions. *Metrika*, 67:63–68, 2008.
- [20] A.W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 2000.
- [21] B. von Bahr and C.-G. Ess  en. Inequalities for the r th absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Annals of Mathematical Statistics*, 36:299–303, 1965.

7 Appendix: some proofs and auxiliary lemmas

We use the following martingale moment inequality.

Lemma 7.1 *Let $p > 1$ and $\{\xi_j, j \geq 1\}$ be a martingale difference sequence: $E[\xi_j | \xi_1, \dots, \xi_{j-1}] = 0$, $j = 2, 3, \dots$ with $E|\xi_j|^p < \infty$. Then there exists a constant $C_p < \infty$ depending only on p and such that*

$$E \left| \sum_{j=1}^{\infty} \xi_j \right|^p \leq C_p \begin{cases} \sum_{j=1}^{\infty} E|\xi_j|^p, & 1 < p \leq 2, \\ (\sum_{j=1}^{\infty} (E|\xi_j|^p)^{2/p})^{p/2}, & p > 2. \end{cases} \quad (7.1)$$

For $1 < p \leq 2$, inequality (7.1) is known as von Bahr and Ess  en inequality, see [21], and for $p > 2$, it is a consequence of the Burkholder and Rosenthal inequality ([3, 16], see also [8, Lemma 2.5.2]).

PROOF OF PROPOSITION 2.1. Since \hat{a}_n in (2.5) is invariant w.r.t. a scale factor of innovations $\{\zeta(t)\}$, w.l.g. we can assume $b^2 + c^2 = 1$ and $E\zeta^2(0) = 1$, $E|\zeta(0)|^{2p} < \infty$. Then $\hat{a}_n - a = \sum_{i=1}^3 \delta_{ni}$, where

$$\begin{aligned} \delta_{n1} &:= -\frac{aX^2(n)}{\sum_{t=1}^n X^2(t) - n(\bar{X}_n)^2}, & \delta_{n2} &:= \frac{\sum_{t=1}^{n-1} X(t)\zeta(t+1)}{\sum_{t=1}^n X^2(t) - n(\bar{X}_n)^2}, \\ \delta_{n3} &:= \frac{\bar{X}_n(X(1) + X(n)) - (\bar{X}_n)^2(1 + n(1-a))}{\sum_{t=1}^n X^2(t) - n(\bar{X}_n)^2}. \end{aligned}$$

The statement of the proposition follows from

$$\Pr(|\delta_{ni}| > \gamma) \leq C(n^{-1} + n^{-(p/2) \wedge (p-1)} \gamma^{-p}) \quad (0 < \gamma < 1, i = 1, 2, 3). \quad (7.2)$$

To show (7.2) for $i = 1$, note that $\delta_{n1} = L_n/(n + D_n)$, where $L_n := -a(1 - a^2)X^2(n)$ and $D_n = D_{n1} - D_{n2}$, $D_{n1} := \sum_{t=1}^n ((1 - a^2)X^2(t) - 1)$, $D_{n2} := n(1 - a^2)(\bar{X}_n)^2$. We have $\Pr(|\delta_{n1}| > \gamma) \leq \Pr(|D_n| > n/2) + \Pr(|L_n| > n\gamma/2)$. Thus, (7.2) for $i = 1$ follows from

$$E|D_{n1}|^{p \wedge 2} \leq Cn, \quad E|D_{n2}| \leq C \quad \text{and} \quad E|L_n|^p \leq C. \quad (7.3)$$

Consider the first relation in (7.3). Clearly, it suffices to prove it for $1 < p \leq 2$ only. We have $D_{n1} = 2D'_{n1} + D''_{n1}$, where

$$\begin{aligned} D'_{n1} &:= (1 - a^2) \sum_{s_2 < s_1 \leq n} \sum_{t=1 \vee s_1}^n a^{2(t-s_1)} a^{s_1-s_2} \zeta(s_1) \zeta(s_2), \\ D''_{n1} &:= (1 - a^2) \sum_{s \leq n} \sum_{t=1 \vee s}^n a^{2(t-s)} (\zeta^2(s) - 1). \end{aligned}$$

We will use the following elementary inequality: for any $-1 \leq a \leq 1, n \geq 1, s \leq n$

$$\begin{aligned} \alpha_n(s) &:= (1 - a^2) \sum_{t=1 \vee s}^n a^{2(t-s)} = \begin{cases} a^{2(1-s)}(1 - a^{2n}), & s \leq 0, \\ 1 - a^{2(n+1-s)}, & 1 \leq s \leq n \end{cases} \\ &\leq C \begin{cases} a^{-2s} \min(1, 2n(1 - |a|)), & s \leq 0, \\ 1, & 1 \leq s \leq n. \end{cases} \end{aligned} \quad (7.4)$$

Using the independence of $\{\zeta(s)\}$ and a and inequality (7.1) (twice) for $1 < p \leq 2$ we obtain

$$\begin{aligned} \mathbb{E}|D'_{n1}|^p &= \mathbb{E} \left| \sum_{s_1 \leq n} \alpha_n(s_1) \zeta(s_1) \sum_{s_2 < s_1} a^{s_1-s_2} \zeta(s_2) \right|^p \\ &\leq C \mathbb{E} \sum_{s_1 \leq n} |\alpha_n(s_1) \zeta(s_1)|^p \sum_{s_2 < s_1} a^{s_1-s_2} \\ &\leq C \mathbb{E} \sum_{s_1 \leq n} |\alpha_n(s_1)|^p \sum_{s_2 < s_1} |a|^{p(s_1-s_2)} \\ &\leq C \mathbb{E}(1 - |a|)^{-1} \sum_{s \leq n} |\alpha_n(s)|^p \leq Cn \end{aligned}$$

since $\mathbb{E}(1 - |a|)^{-1} < \infty$ (see (2.3)) and $\sum_{s \leq n} |\alpha_n(s)|^p \leq Cn$ follows from (7.4). Similarly, since $\{\zeta^2(s) - 1, s \leq n\}$ form a martingale difference sequence,

$$\mathbb{E}|D''_{n1}|^p \leq C \mathbb{E} \sum_{s \leq n} |\alpha_n(s)|^p \leq Cn,$$

proving the first inequality (7.3). The second inequality in (7.3) follows by noting that $n\bar{X}_n = \sum_{s \leq n} (\sum_{t=1 \vee s}^n a^{t-s}) \zeta(s)$ and

$$(1 - a^2) \mathbb{E}[(n\bar{X}_n)^2 | a] = a^2 \left(\frac{1 - a^n}{1 - a} \right)^2 + (1 - a^2) \sum_{s=1}^n \left(\frac{1 - a^s}{1 - a} \right)^2 \leq \frac{Cn}{1 - a}.$$

Consider the last inequality in (7.3). We have $|L_n| \leq |2L'_n + L''_n + 1|$, where

$$L'_n := (1 - a^2) \sum_{s_2 < s_1 \leq n} a^{2(n-s_1)} a^{s_1-s_2} \zeta(s_1) \zeta(s_2), \quad L''_n := (1 - a^2) \sum_{s \leq n} a^{2(n-s)} (\zeta^2(s) - 1).$$

We use Lemma 7.1, as above. Let $1 \leq p \leq 2$. Then $\mathbb{E}|L''_n|^p \leq C \mathbb{E} \sum_{s \leq n} \{(1 - a^2) a^{-2(n-s)}\}^p \leq C$ and $\mathbb{E}|L'_n|^p \leq C \mathbb{E} \sum_{s_2 < s_1 \leq n} \{(1 - a^2) |a|^{2(n-s_1)} |a|^{s_1-s_2}\}^p \leq C \mathbb{E}(1 - |a|)^{p-2} \leq C$. Next, let $p \geq 2$. Then $\mathbb{E}|L''_n|^p \leq C \mathbb{E} \{\sum_{s \leq n} |(1 - a^2) a^{2(n-s)}|^2\}^{p/2} \leq C$ and $\mathbb{E}|L'_n|^p \leq C \mathbb{E}(1 - a^2)^p \{\sum_{s_2 < s_1 \leq n} a^{4(n-s_1)} a^{2(s_1-s_2)}\}^{p/2} \leq C$, proving (7.3) and hence (7.2) for $i = 1$.

Consider (7.2) for $i = 2$. We have $\delta_{n2} = R_n/(n + D_n)$, where $R_n := (1 - a^2) \sum_{t=1}^{n-1} X(t) \zeta(t+1)$ and D_n is the same as in (7.3). Then $\Pr(|\delta_{n2}| > \gamma) \leq \Pr(|R_n| > n\gamma/2) + \Pr(|D_n| > n/2)$, where

$$\begin{aligned} \Pr(|D_n| > n/2) &\leq (n/4)^{-(p \wedge 2)} \mathbb{E}|D_{n1}|^{p \wedge 2} + (n/4)^{-1} \mathbb{E}|D_{n2}| \\ &\leq C \begin{cases} n^{-(p-1)}, & 1 < p \leq 2, \\ n^{-1}, & p > 2, \end{cases} \end{aligned} \quad (7.5)$$

according to (7.3). Therefore (7.2) for $i = 2$ follows from

$$\mathbb{E}|R_n|^p \leq C \begin{cases} n, & 1 < p \leq 2, \\ n^{p/2}, & p > 2. \end{cases} \quad (7.6)$$

Since $R_n = (1 - a^2) \sum_{s \leq n-1} \zeta(s) \sum_{t=1 \vee s}^{n-1} a^{t-s} \zeta(t+1)$ is a sum of martingale differences, by inequality (7.1) with $1 < p \leq 2$ we obtain

$$\begin{aligned}
\mathbb{E}|R_n|^p &\leq CE \sum_{s \leq n-1} |(1 - a^2)\zeta(s) \sum_{t=1 \vee s}^{n-1} a^{t-s} \zeta(t+1)|^p \\
&\leq CE|1 - a^2|^p \sum_{s \leq n-1} \sum_{t=1 \vee s}^{n-1} |a|^{p(t-s)} \\
&\leq CE|1 - a^2|^p \left(\sum_{s \leq 0} |a|^{-ps} \sum_{t=1}^{n-1} |a|^{pt} + \sum_{s=1}^{n-1} \sum_{t=s}^{n-1} |a|^{p(t-s)} \right) \\
&\leq CE|1 - a^2|^p \{ (1 - |a|^p)^{-2} + n(1 - |a|^p)^{-1} \} \leq Cn,
\end{aligned}$$

proving (7.6) for $p \leq 2$. Similarly, using (7.1) with $p > 2$ we get

$$\begin{aligned}
\mathbb{E}|R_n|^p &= \mathbb{E} \left[|1 - a^2|^p \mathbb{E} \left[\left| \sum_{s \leq n-1} \zeta(s) \sum_{t=1 \vee s}^{n-1} a^{t-s} \zeta(t+1) \right|^p |a \right] \right] \\
&\leq CE \left[|1 - a^2|^p \left\{ \sum_{s \leq n-1} \left(\mathbb{E} \left[\left| \zeta(s) \sum_{t=1 \vee s}^{n-1} a^{t-s} \zeta(t+1) \right|^p |a \right] \right)^{2/p} \right\}^{p/2} \right] \\
&\leq CE|1 - a^2|^p \left\{ \sum_{s \leq n-1} \sum_{t=1 \vee s}^{n-1} a^{2(t-s)} \right\}^{p/2} \\
&\leq CE|1 - a^2|^p \left\{ \sum_{s \leq 0} a^{-2s} \sum_{t=1}^{n-1} a^{2t} + \sum_{s=1}^{n-1} \sum_{t=s}^{n-1} a^{2(t-s)} \right\}^{p/2} \\
&\leq CE|1 - a^2|^p \{ (1 - a^2)^{-2} + n(1 - a^2)^{-1} \}^{p/2} \leq Cn^{p/2},
\end{aligned}$$

proving (7.6) and (7.2) for $i = 2$.

It remains to prove (7.2) for $i = 3$. Similarly as above, $\Pr(|\delta_{n3}| > \gamma) \leq \Pr(|Q_n| > n\gamma/2) + \Pr(|D_n| > n/2)$, where $Q_n := (1 - a^2)\{\bar{X}_n(X(1) + X(n)) - (\bar{X}_n)^2(1 + n(1 - a))\}$ and D_n is evaluated in (7.5). Thus, (7.2) for $i = 3$ follows from (7.5) and

$$\mathbb{E}|Q_n|^p \leq C\{\mathbb{E}|(1 - a^2)X^2(n)|^p + \mathbb{E}|(1 - a^2)(\bar{X}_n)^2|^p + n^p \mathbb{E}|(1 - a)(1 - a^2)(\bar{X}_n)^2|^p\} \leq C. \quad (7.7)$$

Since $n\bar{X}_n = \sum_{s \leq n} (\sum_{t=1 \vee s}^n a^{t-s})\zeta(s)$, an application of the second inequality of (7.1) yields

$$\mathbb{E}[|n\bar{X}_n|^{2p}|a] \leq C \left(\frac{(1 - a^n)^2}{(1 - a^2)(1 - a)^2} + \sum_{s=1}^n \left(\frac{1 - a^s}{1 - a} \right)^2 \right)^p.$$

Using $1 - a^n \leq 1 \wedge (n(1 - a))$ we obtain $\mathbb{E}|(1 - a)(1 - a^2)(\bar{X}_n)^2|^p \leq Cn^{-p}$ and $\mathbb{E}|(1 - a^2)(\bar{X}_n)^2|^p \leq CE(1 - a)^{-1}n^{-1}$. Finally, $\mathbb{E}|(1 - a^2)X^2(n)|^p \leq C$ follows by the same arguments as $\mathbb{E}|L_n|^p \leq C$ (see (7.3)). This proves (7.7), thereby completing the proof of (7.2) and of the proposition, too. \square

Let a, a_1, \dots, a_N be i.i.d. r.v.s with d.f. $G(x) = \Pr(a \leq x)$ supported by $[-1, 1]$. Define $\hat{G}_N(x) := N^{-1} \sum_{i=1}^N \mathbf{1}(a_i \leq x)$, $U_N(x) := N^{1/2}(\hat{G}_N(x) - G(x))$, $x \in \mathbb{R}$, and $\omega_N(\delta)$ (= the modulus of continuity of U_N) by

$$\omega_N(\delta) := \sup_{0 \leq y-x \leq \delta} |U_N(y) - U_N(x)|, \quad \delta > 0.$$

Lemma 7.2 Assume that G satisfies Assumption A_6 . Then for all $\epsilon > 0$,

$$\epsilon^4 \Pr(\omega_N(\delta) > 6\epsilon) \leq (3 + 3C)L_G\delta^2 + N^{-1},$$

where C is a constant independent of ϵ, δ, N .

PROOF. As in [2, p. 106, (13.17)] we have that

$$\begin{aligned} \mathbb{E}|U_N(y) - U_N(x)|^2 |U_N(z) - U_N(y)|^2 &\leq 3 \Pr(a \in (x, y]) \Pr(a \in (y, z]), \\ \mathbb{E}|U_N(y) - U_N(x)|^4 &\leq 3 \Pr(a \in (x, y])^2 + N^{-1} \Pr(a \in (x, y]) \end{aligned}$$

for $-1 \leq x \leq y \leq z \leq 1$, where the second inequality treats the 4th central moment of a binomial variable. Now fix $\delta > 0$ and split $[-1, 1] = \cup_i \Delta_i$, where $\Delta_i = [-1 + i\delta, -1 + (i+1)\delta]$, $i = 0, 1, \dots, \lfloor 2/\delta \rfloor - 1$, $\Delta_{\lfloor 2/\delta \rfloor} = [-1 + \lfloor 2/\delta \rfloor \delta, 1]$. According to [17, p. 49, Lemma 1], for all $\epsilon > 0$,

$$\epsilon^4 \Pr(\omega_N(\delta) > 6\epsilon) \leq (3 + 3C) \max_i \Pr(a \in \Delta_i) + N^{-1},$$

where C is a constant independent of ϵ, δ, N . Lemma follows from Assumption A_6 on the d.f. G of the r.v. a . \square

Note that if we take $\delta = \delta_N = o(1)$, we then get $\Pr(\omega_N(\delta) > \epsilon) \rightarrow 0$ as $N \rightarrow \infty$.

Lemma 7.3 Let $\hat{a}_{1,n}, \hat{a}_{2,n}$ be given in (3.3) under Assumptions A_1 – A_6 with $\varrho = 1$. Then for all $\gamma \in (0, 1)$ and $n \geq 1$, it holds

$$\sup_{x, y \in [-1, 1]} |\Pr(\hat{a}_{1,n} \leq x, \hat{a}_{2,n} \leq y) - \Pr(\hat{a}_{1,n} \leq x) \Pr(\hat{a}_{2,n} \leq y)| = O(n^{-(p/2) \wedge (p-1)/(1+p)}).$$

PROOF. Define $\delta_{i,n} := \hat{a}_{i,n} - a_i$, $i = 1, 2$. For $\gamma \in (0, 1)$, we have

$$\Pr(|\delta_{1,n}| > \gamma \text{ or } |\delta_{2,n}| > \gamma) \leq \Pr(|\delta_{1,n}| > \gamma) + \Pr(|\delta_{2,n}| > \gamma) \leq C(n^{-(p/2) \wedge (p-1)})\gamma^{-p} + n^{-1}$$

by Proposition 2.1. Consider now

$$\begin{aligned} \Pr(\hat{a}_{1,n} \leq x, \hat{a}_{2,n} \leq y) &= \Pr(a_1 + \delta_{1,n} \leq x, a_2 + \delta_{2,n} \leq y) \\ &\leq \Pr(a_1 + \delta_{1,n} \leq x, a_2 + \delta_{2,n} \leq y, |\delta_{1,n}| \leq \gamma, |\delta_{2,n}| \leq \gamma) + \Pr(|\delta_{1,n}| > \gamma \text{ or } |\delta_{2,n}| > \gamma). \end{aligned}$$

Then

$$\begin{aligned} \Pr(a_1 + \delta_{1,n} \leq x, a_2 + \delta_{2,n} \leq y, |\delta_{1,n}| \leq \gamma, |\delta_{2,n}| \leq \gamma) &\leq \Pr(a_1 \leq x + \gamma, a_2 \leq y + \gamma, |\delta_{1,n}| \leq \gamma, |\delta_{2,n}| \leq \gamma) \\ &\leq G(x + \gamma)G(y + \gamma) \end{aligned}$$

and

$$\begin{aligned} \Pr(a_1 + \delta_{1,n} \leq x, a_2 + \delta_{2,n} \leq y, |\delta_{1,n}| \leq \gamma, |\delta_{2,n}| \leq \gamma) &\geq \Pr(a_1 \leq x - \gamma, a_2 \leq y - \gamma, |\delta_{1,n}| \leq \gamma, |\delta_{2,n}| \leq \gamma) \\ &\geq G(x - \gamma)G(y - \gamma) - \Pr(|\delta_{1,n}| > \gamma \text{ or } |\delta_{2,n}| > \gamma). \end{aligned}$$

From (2.6) we obtain

$$|G(x \pm \gamma)G(y \pm \gamma) - G(x)G(y)| = |(G(x) + O(\gamma))(G(y) + O(\gamma)) - G(x)G(y)| \leq C\gamma.$$

Hence,

$$|\Pr(a_1 \leq x, a_2 \leq y) - G(x)G(y)| \leq C(\gamma + n^{-1} + n^{-(p/2) \wedge (p-1)} \gamma^{-p}). \quad (7.8)$$

In a similar way,

$$|\Pr(a_1 \leq x) \Pr(a_2 \leq y) - G(x)G(y)| \leq C(\gamma + n^{-1} + n^{-(p/2) \wedge (p-1)} \gamma^{-p}). \quad (7.9)$$

By (7.8), (7.9), the proof of the lemma is complete with $\gamma = \gamma_n = o(1)$, which satisfies $\gamma_n \sim n^{-(p/2) \wedge (p-1)} \gamma_n^{-p}$.

□